

Cleveland State University
EngagedScholarship@CSU



Electrical Engineering & Computer Science Faculty
Publications

Electrical Engineering & Computer Science
Department

3-2006

Selective Advance Reservations Based on Host Movement Detection and Resource-Aware Handoff

Kyounghee Lee

Information and Communications University, leekhe@icu.ac.kr

Myungchul Kim

Information and Communications University, mckim@icu.ac.kr

Chansu Yu

Cleveland State University, c.yu91@csuohio.edu

Ben Lee

Follow this and additional works at: https://engagedscholarship.csuohio.edu/enece_facpub

benlee@ecs.uoregonstate.edu

 Part of the [Systems and Communications Commons](#)

Seungphil Hong

How does access to this work benefit you? Let us know!

LG CNS, philhong@lgcns.com

Publisher's Statement

This is the accepted version of the following article: Lee, K., Kim, M., Yu, C., Lee, B., , & Hong, S. (2006). Selective advance reservations based on host movement detection and resource-aware handoff. *International Journal of Communication Systems*, 19(2), 163 - 184. doi:10.1002/dac.779, which has been published in final form at <http://onlinelibrary.wiley.com/doi/10.1002/dac.779/abstract>

Original Citation

Lee, K., Kim, M., Yu, C., Lee, B., , & Hong, S. (2006). Selective advance reservations based on host movement detection and resource-aware handoff. *International Journal of Communication Systems*, 19(2), 163 - 184. doi:10.1002/dac.779

Repository Citation

Lee, Kyounghee; Kim, Myungchul; Yu, Chansu; Lee, Ben; and Hong, Seungphil, "Selective Advance Reservations Based on Host Movement Detection and Resource-Aware Handoff" (2006). *Electrical Engineering & Computer Science Faculty Publications*. S2.

https://engagedscholarship.csuohio.edu/enece_facpub/S2

This Article is brought to you for free and open access by the Electrical Engineering & Computer Science Department at EngagedScholarship@CSU. It has been accepted for inclusion in Electrical Engineering & Computer Science Faculty Publications by an authorized administrator of EngagedScholarship@CSU. For more information, please contact library.es@csuohio.edu.

Selective advance reservations based on host movement detection and resource-aware handoff

Kyounghee Lee^{1,*†}, Myungchul Kim^{1,‡}, Chansu Yu^{2,§}, Ben Lee^{3,¶}
and Seungphil Hong^{4,||←}

¹ *School of Engineering, Information and Communications University, Daejeon, Korea*

² *Department of Electrical and Computer Engineering, Cleveland State University, Cleveland, OH, U.S.A.*

³ *School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR, U.S.A.*

⁴ *SNS Team, Infra Solution Division, LG CNS, Seoul, Korea*

SUMMARY

This paper proposes a new mechanism, which addresses the excessive advance reservation requirements of QoS guarantee methods for mobile Internet. To save resources for excessive advance reservations, the proposed mechanism employs a movement detection scheme for a mobile host (MH) using link-layer functionalities. With the movement detection scheme, advance reservations can be established at only where a MH is likely to visit soon. Another novel feature of our mechanism is resource-aware handoff direction scheme that allows a MH to choose its next BS according to not only the link-layer signal strength, but also the available amount of resources in the reachable base stations (BSs). It considerably decreases a probability that QoS is disrupted due to the failure in advance reservation request. Also, the proposed mechanism requires fewer functional and structural changes to the current Internet components and protocols since all the enhanced features are integrated only into leaf BSs and MHs. It does not suffer from the problems of the conventional approaches based on Mobile IP and RSVP Tunnel, such as non-optimal routing path and signalling overhead. Our experiment results show that the proposed mechanism successfully eliminates the overhead for useless advance reservations while guaranteeing seamless QoS for MHs. The performance comparison demonstrates that our mechanism slightly outperforms the conventional approaches while requiring fewer modifications and additions to the existing Internet architecture. This performance advantage of the proposed mechanism becomes noticeable when the network is congested and the mobility of a host is high.

KEY WORDS: QoS; mobile Internet; advance reservations; movement detection; handoff direction

*Correspondence to: Kyounghee Lee, School of Engineering, Information and Communications University, Daejeon, Korea.

†E-mail: leekhe@icu.ac.kr

‡E-mail: mckim@icu.ac.kr

§E-mail: c.yu91@csuohio.edu

¶E-mail: benl@ece.orst.edu

||E-mail: philhong@lgcns.com

1. INTRODUCTION

As mobile devices have more computing power and wireless links provide higher bandwidth, multimedia-streaming services are becoming one of the most attractive and challenging applications in the mobile Internet. However, there remain some limitations in the current mobile Internet that make such services difficult to deploy, including service disruption due to handover latency and traffic path redirection overhead as well as poor communication characteristics in wireless networks. Therefore, one of the ultimate challenges for mobile multimedia-streaming service is to provide continuous QoS guarantees while a mobile host (MH) moves across multiple wireless cells.

Several useful mechanisms including Resource reSerVation Protocol (RSVP) [1] and Differentiated Services (DiffServ) architecture [2] have been proposed for QoS provisioning in wired Internet. RSVP is a signalling protocol that facilitates QoS guarantees by allowing a destination node to reserve resources along a fixed path to a source node. Even though it guarantees the desired QoS in the wired Internet, there are two major reasons that make RSVP inapplicable to Mobile IP [3] networks: First, the IP-in-IP encapsulation scheme of Mobile IP makes RSVP signal messages invisible to the intermediate routers in IP tunnels. Second, the previously reserved resources no longer become valid after a host moves to a new cell, which results in additional overhead and latency to establish a new resource reservation path.

There have been a number of approaches [4–10] to address the aforementioned problems. Mobile RSVP (MRSVP) [5,6] introduced an advance resource reservation, called *passive reservation*, to adapt RSVP to the mobile Internet. A drawback of the approach is that the excessive passive reservations waste network resources. Thus, more recent approaches [7–10] focus mainly on reducing the overhead and delay caused by the advance resource reservation scheme. However, in most conventional approaches, all neighbouring cells of a MH should have their own advance reservations even though the MH will only visit one of them. Only Hierarchical MRSVP (HMRSVP) [10] addresses this issue, but it requires considerable modifications to the existing Internet protocols and components to support RSVP Tunnel [4] and Mobile IP regional registration [11]. Another drawback of HMRSVP is to require additional signalling overhead to maintain multiple RSVP tunnels for a single reservation session.

This paper proposes a new mechanism, called *Selective Advance Reservations and resource-Aware Handoff direction (SARAH)*, to guarantee seamless QoS for mobile Internet with RSVP. To support host mobility, SARAH dynamically extends and reduces a reservation path for a MH by using an advance reservation, called a *pseudo reservation* in this mechanism. The proposed mechanism provides a number of architectural advantages over the existing approaches. First, SARAH employs a link-layer movement detection scheme to predict a MH's next base station (BS). This saves network resources and improves QoS guarantees by avoiding multiple useless advance reservations. Second, SARAH has a novel feature to utilize network resources more efficiently, called a resource-aware handoff direction scheme. With this scheme, a MH can choose its next BS based on not only the signal strength of link-layer control frames, but also the available resources at the adjacent BSs. This scheme is very effective in improving QoS guarantees attained by SARAH, particularly in congested networks. Third, SARAH guarantees that the establishment of pseudo reservations always finishes before the completion of the Mobile IP handoff. This assures that SARAH imposes no additional service disruption on the Mobile IP handoff latency. Finally, SARAH requires fewer functional and

structural changes to the current Internet environment, and more importantly, requires no changes to the existing RSVP and Mobile IP protocol. Only leaf BSs and MHs are equipped with the enhanced features to create pseudo reservations and to perform the reservation path extension procedure. Since a pseudo reservation is established only between two neighbouring leaf BSs, SARAH shortens the average length of an advance reservation path and enables all network components except leaf BSs to handle the pseudo reservations with no functional and structural changes.

The performance of SARAH is analysed using a real system and compared with other existing approaches based on simulations. First, a testbed was built to show the operability of SARAH. The measured results from our implementation illustrate that SARAH adds no further delays to the original Mobile IP handoff latency and continuously guarantees a desirable QoS when a MH moves to a congested network. Second, the simulation results of the proposed mechanism are compared to two existing methods: MRSVP [5] and HMRSVP [10]. The comparison shows that SARAH outperforms MRSVP and provides similar QoS guarantees with HMRSVP in terms of the reservation session completion rate, which represents the probability that a MH can finish a reservation session without suffering from any QoS degradation. The performance advantage of SARAH compared to HMRSVP becomes noticeable as the offered load in the network becomes high and the average number of handoffs increases during a reservation session, i.e. as the network becomes congested and the host mobility increases. This illustrates that SARAH is an efficient way to guarantee QoS in the congested mobile access networks without support of intermediate routers.

The rest of the paper is organized as follows: Section 2 discusses related work on RSVP with mobility support. Section 3 presents an overview of the proposed SARAH mechanism. Section 4 describes our movement detection scheme, resource-aware handoff direction scheme, and how SARAH extends and reduces a reservation path to support host mobility. Sections 5 and 6 illustrate our experimental testbed and the simulation model, and present the obtained results from the experiments. Finally, Section 7 concludes the paper.

2. RELATED WORK

A lot of effort has been devoted to adapt RSVP to Mobile IP networks. One of the earliest approaches is RSVP Tunnel [4] scheme. This scheme addressed the RSVP message invisibility problem within an IP tunnel but it did not consider the problem that a reservation path becomes invalid after a MH changes its location. Thus, with RSVP Tunnel, a MH cannot escape temporary QoS degradation after a handoff. Since RSVP Tunnel is based on Mobile IP, all the data destined to a MH should pass the MH's home agent (HA) [3]. This non-optimal routing path requires more time and network resources to establish a RSVP session for a MH. Another drawback of RSVP Tunnel is additional signalling overhead at a MH's mobile agents (MAs).

Some later works employed an advance reservation to prepare for a MH's possible movement. In MRSVP [5, 6], an advance reservation, called a *passive reservation*, was introduced by Talukdar *et al.* A passive reservation reserves resources in advance at a neighbour of the current cell, but no traffic is delivered on it until it is activated after the MH's handoff. With those passive reservations, the QoS degradation due to establishment of a new RSVP session after a handoff can be eliminated. In this scheme, a special host, called *proxy agent*, performs active/passive reservations on behalf of the MH. Since the passive reservations are

established along a multicast tree consisting of a correspondent host (CH) and all the proxy agents in the neighbouring cells, the overhead and resource consumption for the excessive passive reservations can be several times higher than for active reservations. Another limitation of the scheme is that the passive reservation functions need to be added to all routers in the network. Mahadevan and Sivalingam [9] proposed a network architecture that requires fewer passive reservation-capable routers than MRSVP. In this approach, all the passive reservations are established between two neighbouring BSs. Thus, when a MH moves to a neighbouring cell, the corresponding passive reservation is activated and attached to the original RSVP path and then the traffic is delivered along the extended reservation path. Though the architecture saves backbone routers from the passive reservation overhead, it still requires a significant number of gateway routers to be equipped with passive reservation capabilities. This is because the approach establishes a passive reservation between the neighbouring BS and its gateway router to manage a handoff from a QoS domain (normally, a routing domain) to another. It is to prevent the infinite extension of a reservation path.

Subsequently, there have been several approaches to reduce the overhead required for advance reservations. Chen and Huang [8] described a method similar to MRSVP, which employs *predictive reservation* and *temporary reservation* schemes. Predictive reservations are advance reservations made for all the locations where a MH may visit. These locations form a group, which is composed of the leaves of a multicast tree, and the movement of a MH is modelled as a membership transition between the multicast groups. In this method, to make more efficient use of network resources, temporary reservations can use the inactive bandwidth that have already been reserved by the other predictive reservations. Tseng *et al.* proposed the Hierarchical MRSVP (HMRSVP) [10] to reduce the overhead due to excessive passive reservations in MRSVP. It saves resources by establishing passive reservations only when a MH moves between two different regions, possibly between two routing domains. To do this, however, considerable modifications are needed to the current Internet protocols and components for supporting RSVP Tunnel [4], Mobile IP regional registration [11], and passive reservation schemes. HMRSVP requires additional signalling overhead to maintain multiple RSVP tunnels for a single reservation session. Moreover, the QoS disruption time for an inter-region handoff can be longer than the Mobile IP handoff time since the establishment of passive reservations starts with the Mobile IP registration with the HA.

Another effort to escape the advance reservation overhead is to maintain a single contact address of a MH during a RSVP session. Pasklis *et al.* [7] introduced a scheme that a RSVP mobility proxy (RSVP-MP) in the access network dynamically updates its own binding between a MH's Local Care-of Address (LCoA) and Domain Care-of Address (DCoA) whenever there is any change in the MH's address. Since a RSVP-MP performs appropriate address translation of RSVP messages and data packets, a MH can always be represented by a single IP address (i.e. DCoA) while it is moving within an access network. However, a RSVP-MP may suffer from the excessive burden of address translation for all packets destined to MHs. This approach requires a special mechanism to maintain a single contact IP address inside a domain. It also requires another QoS technology, such as DiffServ, to support wider mobility between different access networks.

In summary, most existing approaches based on advance reservation do not address the problem that all neighbouring locations should have their own advance reservations but only one of them will be actually used by a MH. Even though HMRSVP [10] provides a possible solution to this problem, it requires considerable modifications to the current Internet

environment and incurs additional signalling overhead for managing multiple RSVP tunnels. Therefore, this paper proposes a new mechanism that efficiently reduces the number of advance reservations while guaranteeing seamless QoS for a MH and requiring minimal changes to the current Internet protocols and components. In the following sections, a detailed description of the proposed mechanism is presented.

3. SARAH OVERVIEW

In SARAH, an advance reservation, called *pseudo reservation*, is used in place of the passive reservation in MRSVP [5, 6]. A pseudo reservation session is established and released using RSVP *path*, *resv*, and *path teardown* messages [1] as the same way as a normal RSVP session. In most existing approaches including [5, 8, 10], advance reservations are established along routing paths between a MH's neighbouring cells and the CH. This requires that a crossover node (i.e. a CH or an intermediate router) of the MH's active reservation path and advance reservation path is equipped with additional functions to create and to handle the advance reservations in a special manner. The additional functions include newly defined RSVP signalling messages for advance reservations, traffic blocking for inactivated advance reservation paths, traffic redirection or forwarding over the advance reservation path after a handoff. However, in SARAH, only two geographically neighbouring BSs can make, maintain and release a pseudo reservation between them. Thus, any network components except leaf BSs do not need to know whether a RSVP session is a pseudo or an active reservation. Since a pseudo reservation is established and handled as the same way as a normal RSVP session, SARAH does not require any modifications on the existing RSVP protocol. To prevent the waste of resources, each BS of a pseudo reservation session does not deliver any traffic over the session until it is activated. The inactive resources allocated to a pseudo reservation can be temporarily used by the other best-effort traffic. The reservation path extension after a handoff is also achieved by two end BSs' traffic forwarding over the pseudo reservation session without requiring any additional features such as RSVP Tunneling [4]. This enables SARAH to integrate all the enhanced features for pseudo reservation and path extension into the leaf BSs.

Figure 1 shows the overall SARAH procedure that consists of three steps: (a) PRP establishment before a handoff, (b) *extension of reservation path* (ERP) process after a handoff, and (c) *optimization for extended reservation path* (ORP) process. In Step (a), with the movement detection scheme, which will be described in the following section, a BS pre-establishes *pseudo reservation paths* (PRPs) only with one or a few predicted neighbouring BSs that a MH is likely to visit in a moment. This detection can be achieved using link-layer control frames when a MH enters into the overlapped area between two adjacent cells. Figure 1(a) illustrates the first step that an inter-routing-domain PRP is established between BS_B and BS_C. When a MH moves to one of the predicted cells and finishes the Mobile IP handoff procedure to the new BS, a reservation path for the MH is immediately extended by the ERP process. As shown in Figure 1(b), Step (b) involves activating a corresponding PRP (i.e. a PRP between the current cell and the previous cell) and attaching it to the original RSVP session. The previous BS (BS_B) then forwards traffic between these two reservation sessions. In the final step, SARAH dynamically terminates useless PRPs after a handoff and optimizes the extended reservation path to avoid the overhead due to data delivery through a non-optimal routing path. As shown

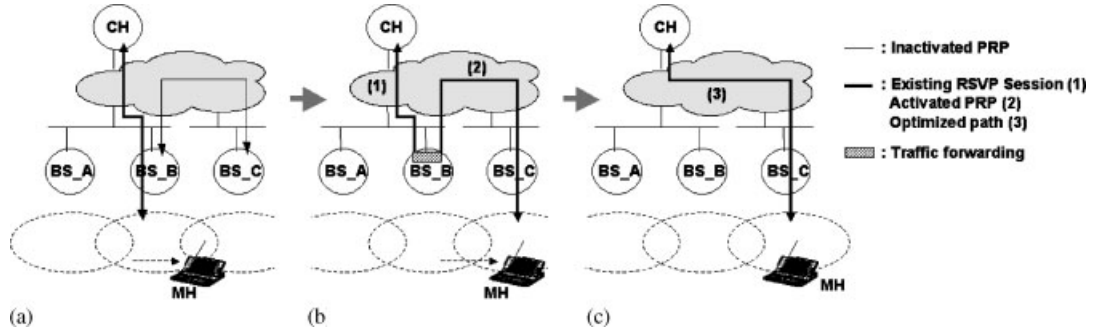


Figure 1. Overall SARAH process: (a) Pseudo reservation; (b) ERP process after a Landoff; and (c) ORP process.

in Figure 1(c), the ORP process is achieved by adjusting the extended reservation path to a general routing path between the CH and MH.

4. RESERVATION PATH EXTENSION AND OPTIMIZATION

In this section, we give a detailed description of the proposed SARAH mechanism. The description includes the host movement detection scheme, the resource-aware handoff direction scheme, and how SARAH extends and optimizes a reservation path to support host mobility.

4.1. Host movement detection

Some approaches [12, 13] have been proposed to reduce the Mobile IP handoff latency by tightly coupling the layer-3 handoff process with the layer-2 (link layer or L2) functionality. A main idea of SARAH movement detection scheme is similar with those approaches in using L2 functionality to predict a MH's next location. Mobile IP was originally designed without any assumptions of the underlying link-layer to provide the widest applicability. However, in SARAH, it is assumed that a MH can simultaneously detect L2 *beacon* frames from multiple wireless attach points (APs). In this paper, such underlying network was built with the IEEE 802.11 Wireless LANs [14]. Another assumption to enable SARAH movement detection scheme is that there should exist an overlapped area between two neighbouring wireless cells. Since Mobile IP registration with a new foreign agent (FA) begins after the completion of L2 roaming procedure, detecting L2 beacons is an effective way to recognize a MH's movement without incurring a Mobile IP handoff delay. With this movement detection scheme, SARAH can predict the next BS of a MH that enters into the overlapped area of the two adjacent wireless cells. This enables SARAH to reduce the number of required PRPs significantly.

In Figure 2, as a MH moves from BS1 to BS2, the signal-to-noise ratio (SNR) of BS1 falls below the *cell search threshold* (CST) value where the MH finds another reachable APs with active scanning [15]. Then the MH can receive L2 beacon frames from not only BS1, but also BS2. This becomes a trigger that SARAH starts the establishment of a PRP between BS1 and the predicted next BS (i.e. BS2 in the figure). When the MH reaches the *cell switchover point*

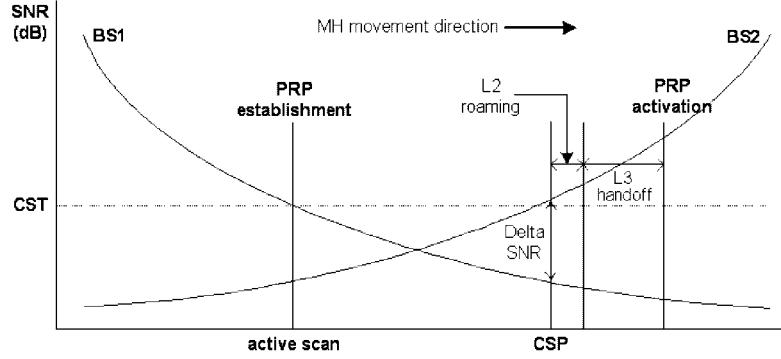


Figure 2. Movement detection for a MH.

(CSP) that the difference between SNR values of BS1 and BS2 exceeds ΔSNR , the L2 roaming procedure is initiated.

To inform the current BS that a MH has come into the overlapped area where L2 beacon frames from multiple BSs can be detected, a notifying message is sent from the MH to the current BS. This message contains the new BS's MAC address obtained from the beacon frames. In SARAH, each BS has a *neighbour mapping table* that binds IP and MAC addresses of all the neighbouring BSs. Thus, the current BS can start the establishment of a PRP with the predicted BS using an IP address in the neighbour mapping table.

Whenever a MH predicts a new wireless cell that it may visit, SARAH establishes a PRP to/from the corresponding BS in the detected cell. However, a movement prediction is performed only when a MH enters into the overlapped region between two or more neighbouring BSs (i.e. the current BS and the predicted BSs). This further reduces the overhead for an advance reservation. As shown in Figure 2, an interval while a PRP is inactivated is from its establishment to the completion of a Mobile IP (L3) handoff. This illustrates that our movement detection scheme does not only reduce the number of establishments of advance reservations, but also significantly shortens a time that resources are wasted by an inactivated advance reservation compared to the existing approaches [5, 8, 9].

4.2. Reservation path extension using PRP

Figure 3 shows the PRP establishment process when a MH is about to leave the current BS (BS_B). To show that SARAH supports the movements between two different routing domains, an example is used where an inter-routing-domain PRP is established. For simplicity, Figure 3 illustrates that the movement detection scheme predicts only one reachable BS (BS_C). If there are multiple predicted BSs, the example in the figure may need to include some additional steps for resource-aware handoff direction. A detailed description of the resource-aware handoff direction scheme will be given in Section 4.3.

As shown in Figure 3(a), when the MH enters the overlapped region between cells of BS_B and BS_C, it is able to receive beacons delivered from BS_C. Then, the MH sends a *PRP_init* message to its current BS (BS_B) to notify the MAC address of BS_C. BS_B searches its neighbour mapping table to obtain the BS_C's IP address, and sends a *PRP_inform* message to

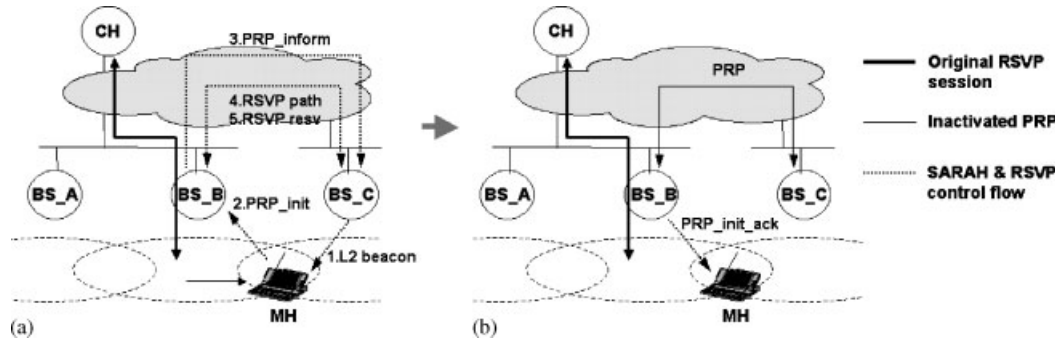


Figure 3. Selective establishment of PRPs.

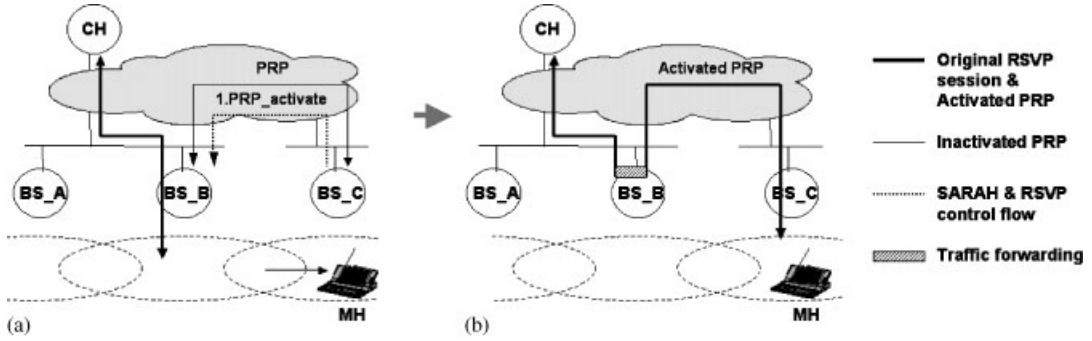


Figure 4. Reservation path extension.

BS_C. This message informs the possibility of the MH's future entrance into the BS_C's cell. A *PRP_inform* message can contain *Tspec*, which defines the traffic characteristics of the data flow delivered over the original RSVP session. These traffic characteristics are used in the RSVP signalling messages, such as *RSVP_path* and *resv* messages, which are exchanged to reserve resources for a PRP between BS_B and BS_C (Figure 3(a)). Finally, a PRP is established as shown in Figure 3(b) and then BS_B notifies the MH of the successful establishment of a PRP by sending a *PRP_init_ack* message.

After a MH completes the Mobile IP handoff procedure to a new BS, the ERP process immediately starts to extend a reservation path to the new location of the MH. Figure 4 shows the ERP process in SARAH. In the proposed architecture, a BS also acts as a MA of Mobile IP. Thus, when a MH visits a new BS, it starts a Mobile IP registration process with its HA via the new BS [3]. By relaying a Mobile IP *registration request* packet from the MH to the HA, the new current BS (BS_C) knows that a PRP should be activated between the previous BS (BS_B) and itself. Therefore, as shown in Figure 4(a), BS_C notifies BS_B of the need of PRP activation by sending a *PRP_activate* message. Then, PRP activation is performed either by the current BS or the previous BS, depending on which one currently acts as the sender of the PRP. Finally, the reservation path is extended by attaching the activated PRP to the original RSVP session. To guarantee seamless QoS to the MH, BS_B forwards the traffic between the activated PRP and the original RSVP session as shown in Figure 4(b).

4.3. Resource-aware handoff direction

SARAH has a novel feature to improve QoS guarantee, the resource-aware handoff direction scheme, which allows a MH to choose its next BS according to not only the signal strength of L2 control frames, but also the available resources in the reachable BSs. In the proposed architecture based on IEEE 802.11 Wireless LAN [14], a MH can determine the candidates for its next BS (i.e. reachable BSs) using L2 *beacon* frames recently delivered from BSs. Otherwise, to determine which BSs are currently reachable, a MH may broadcast a *probe request* frame to trigger *probe response* frames from BSs. When multiple BSs are reachable, a MH associates itself with the BS that delivers the strongest *beacon* or *probe response* frames [15]. However, the resource-aware handoff direction scheme allows a MH to choose one of the several candidate BSs according to the amount of available resources. This increases the probability that a MH completes its reservation session without any QoS degradation due to lack of network resources.

Figure 5 shows the steps of PRP establishment when the resource-aware handoff direction scheme is applied. In Figure 5(a), the MH is currently being served by BS_C and it is about to leave the current cell. With the movement detection scheme in Section 4.1, the MH knows that BS_A and BS_B are the candidates for its next BS since both Beacon_A and Beacon_B are being delivered simultaneously. Assuming that Beacon_A is stronger than Beacon_B, the MH initiates the PRP establishment process by sending a *PRP_init* message containing the MAC address of BS_A when a normal L2 roaming policy is applied. However, when the resource-aware handoff direction scheme is applied, the *PRP_init* message contains a list of MAC addresses of all the reachable BSs (i.e. BS_A and BS_B) in order of their signal strength. Then, based on this order, BS_C first tries to establish a PRP between BS_A and itself.

Step (2) in Figure 5(b) illustrates the PRP establishment process described in Section 4.2. If a PRP is successfully established by Step (2), BS_C determines it does not need to establish additional PRPs for the MH. Thus, Step (3) is skipped and BS_C immediately sends a *PRP_init_ack* message to the MH to notify the completion of PRP establishment (see Step (4)). However, if the PRP request in Step (2) is rejected due to lack of resources at BS_A or the intermediate routers, BS_C tries to find another reachable BS (i.e. BS_B) according to the MAC address order indicated in the *PRP_init* message. Thus, Step (3) is performed to establish a PRP

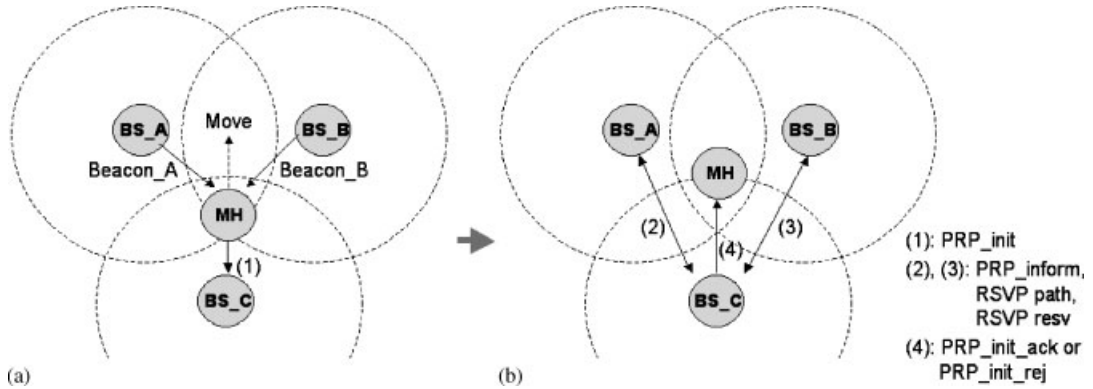


Figure 5. Resource-aware handoff direction scheme.

between BS_B and BS_C. If Step (3) is successful, BS_C sends a *PRP_init_ack* message to the MH as shown in Step (4). A *PRP_init_ack* message is sent only once when a PRP is successfully established for the first time. The message contains MAC and IP addresses of the BS that the previous PRP request has been accepted. This directs the MH to choose its next BS according to whether or not a reachable BS can provide the desirable resources. If all the PRP requests are rejected, BS_C sends a *PRP_init_rej* message to the MH to notify the failure in establishing a PRP. Then, the MH performs a general L2 roaming procedure depending on the signal strength of control frames.

4.4. Optimization for extended reservation path

As a MH continuously moves across the wireless cells, consecutive ERP processes can extend a reservation path too long. Therefore, when necessary, SARAH performs the ORP process to reduce the overhead and waste of resources due to an extended reservation path. The ORP process adjusts the extended reservation path to the optimized one, which is a RSVP session established along the shortest routing path between a sender and a receiver. To determine when an optimization process needs to be performed is an important issue related to both resource utilization and signalling overhead. In the aspect of resource efficiency, we can define some cases that the extended reservation paths should be optimized, such as a reservation path including an inter-routing-domain PRP or a loop. Frequent ORP processes save network resources but may increase the signalling and processing overhead on BSs.

The ORP process can be performed by using either multicast IP address or unicast IP address. The ORP process using unicast IP address establishes a new RSVP session between the current BS and the CH and then replaces the extended reservation path with this new RSVP session. For better network resource utilization, a more efficient way of the ORP process is to use multicast IP address. As illustrated in Figure 6, it reduces resource consumption in the optimization process by joining the existing multicast RSVP session instead of making a new RSVP session.

In Figure 6, it is assumed that a RSVP session with multicast address had been established between the CH and BS_A, and an ERP process has been performed as described before. To initiate an ORP process, BS_A first sends a *PRP_opt* message to BS_B as shown in

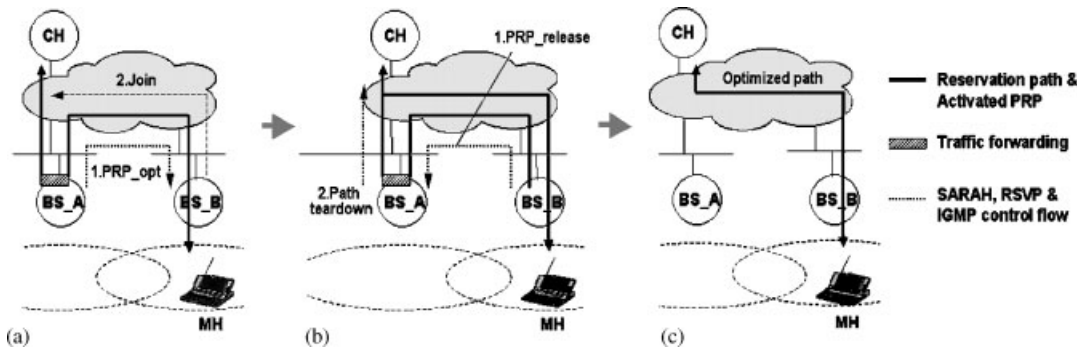


Figure 6. Optimization of reservation path (using multicast IP address).

Figure 6(a). This message contains a multicast group address used by the original RSVP session between CH and BS_A. With the multicast address, BS_B can join the existing multicast RSVP session to acquire a direct reservation path along the shortest path between the CH and itself. Consequently, BS_B is now able to use the new RSVP path to deliver traffic for the MH as shown in Figure 6(b). Then, the needless PRP between BS_A and BS_B is terminated using a *PRP_release* message. After receiving the *PRP_release* message, BS_A leaves the multicast group by sending a *RSVP_path teardown* message. Finally, only the optimized path remains between BS_B and the CH as shown in Figure 6(c).

In the joining process of Figure 6(a), there are some differences depending on whether a MH is a sender or a receiver in the existing RSVP session. When the MH is a sender, BS_B can join in the existing RSVP session by sending a *RSVP_path* message destined to a multicast address of the RSVP session. Subsequently, a receiver of the *RSVP_path* message (i.e. CH) replies with a *RSVP_resv* message allowing BS_B to join the existing RSVP session as a sender. However, when the MH is a receiver, BS_B first joins the IP multicast group using the Internet Group Management Protocol (IGMP) *report* message [16]. Then, it waits for a *RSVP_path* message, which the sender (CH) periodically transmits through the IP multicast session to identify a new destination in the flow [1]. While waiting for a *RSVP_path* message, BS_B should deliver traffic from the activated PRP to the MH to support seamless QoS guarantee. When BS_B receives a *RSVP_path* message, it replies with a *RSVP_resv* message to join the existing multicast RSVP session.

The ORP process using multicast address decreases the probability that an optimization request is rejected by intermediate routers. However, the ORP process using unicast address is required to support cases that the underlying networks do not support IP multicasting or a newly entering MH has already participated in a unicast RSVP session.

5. IMPLEMENTATION

In order to evaluate the performance of the proposed SARAH mechanism, an experimental testbed was implemented to show its practicality and to obtain actual measurements. The architecture of the experimental testbed is shown in Figure 7. Each BS is equipped with SARAH module, Mobile IP foreign agent (FA) module, RSVP signalling module, and routing/traffic-scheduling module. The SARAH module handles all the control messages for ERP/ORP processes. It also directs the RSVP signalling module to perform the required RSVP activities on behalf of a MH. The routing/traffic-scheduling module delivers traffic between the wired and wireless networks, and is responsible for handling incoming/outgoing packets according to the pre-specified scheduling policy.

For wireless communications, each BS and MH is equipped with a WaveLAN card that provides a band-width of 11 Mbps [17]. A gateway router runs the *alternate queueing* (ALTQ) [18] package for traffic scheduling. To support host mobility, the *Dynamics* Mobile IP software [19] was deployed. The RSVP package from University of Southern California [20] was modified to perform RSVP signalling required by the SARAH process.

Using the testbed, we evaluated the performance of SARAH in terms of PRP establishing time, service disruption time after a handoff, and data transmission rate. Figure 8 shows the general procedure to support a MH's handoff and each step's latency based on our testbed configuration. When a MH starts active scanning [15] and then receives L2 beacon frames not

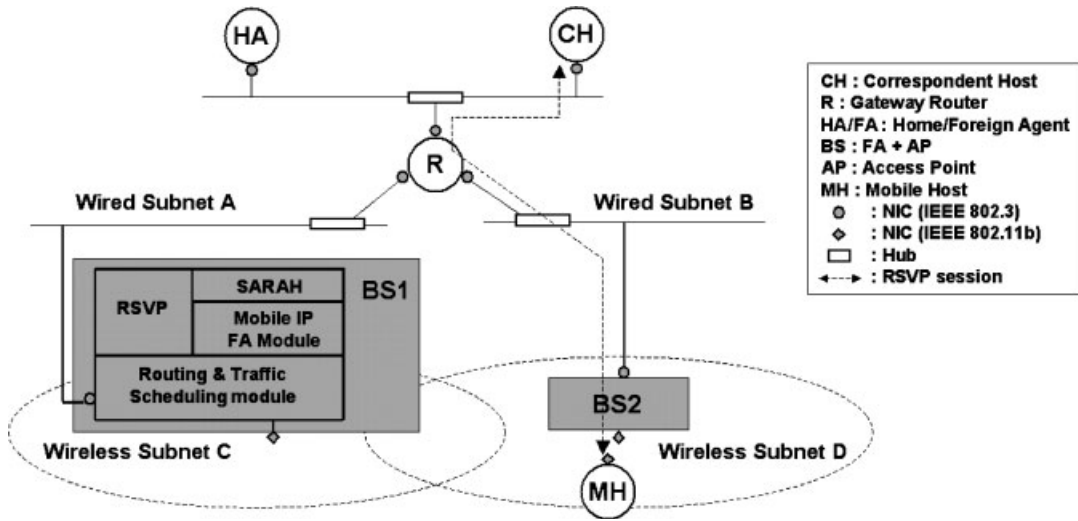


Figure 7. The experimental testbed architecture.

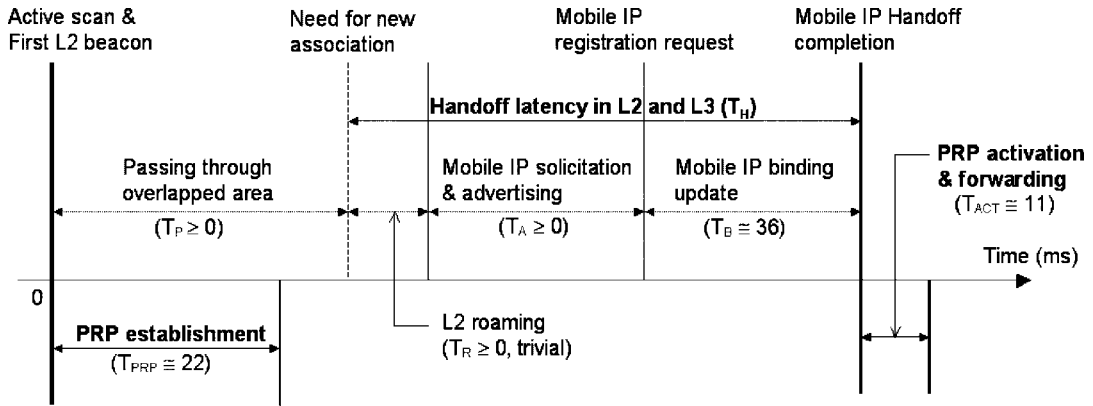


Figure 8. Analysis of handoff latency in Mobile IP and SARAH.

sent by the current BS, the MH knows that it has entered the overlapped area of two wireless cells. Thereafter, the MH can determine when it should associate itself with a new BS by comparing the strength of their beacon frames. The interval (T_P) after receiving a beacon frame from a new BS for the first time and before starting an association with the BS depends on the configuration of wireless networks. That is, T_P varies with the size of overlapped area and the moving speed of the MH. With most implementations of IEEE 802.11 Wireless LANs, the average latency for L2 roaming (T_R) is estimated to be a few milliseconds or less. This latency is generally required for exchanging *association request* and *response* frames [14] between the MH and AP (i.e. a BS in the proposed mechanism).

After the completion of L2 roaming process, a MH waits for an *agent advertisement* message from a new MA or broadcasts an *agent solicitation* message in order to perform the Mobile IP

(L3) handoff procedure [3]. This delay, which is denoted as T_A , is dependent on the pre-configuration of advertisement interval in each MA. If a MA generates an *agent advertisement* every 100 ms, the average of T_A can be estimated to be about 50 ms. Finally, after receiving an *agent advertisement* message, a MH starts the Mobile IP registration procedure by sending a *registration request* message to its HA. We measured the Mobile IP binding update time that represents the period from when a *registration request* message is sent to when the registration reply message is received from HA. The binding update time (T_B) in our testbed is around 36 ms as shown in Figure 8.

On the other hand, the time to establish a PRP (T_{PRP}) in our implementation is on the average about 22 ms when the two neighbouring BSs are 2 hops away from each other as illustrated in Figure 7. Note that the actual delay for the L2 roaming and Mobile IP handoff (T_H) is greater than the MIP binding update time T_B (i.e. 36 ms) in Figure 8. Therefore, the PRP establishing procedure in our implementation is guaranteed to finish before the Mobile IP handoff completes. This means that the PRP establishment does not increase the service disruption of the original Mobile IP handoff latency. If there exist multiple reachable BSs in the testbed, T_{PRP} may increase due to the resource-aware handoff direction scheme. However, the additional overhead is negligible when compared to T_P (see Figure 8). The PRP establishment procedure starts immediately after a MH receives an L2 *beacon* frame from a new BS. Since T_P is quite large in the real networks, the PRP establishment procedure usually completes before the start of L2 roaming. For example, if the width of the overlapped area is 60 m and a MH moves at a speed of 15 m/s, T_P can be estimated to be about 2 seconds or more.

The only noticeable service disruption time in the ERP process is the PRP activation time T_{ACT} , which was measured to be around 11 ms (see Figure 8). Note that the total service disruption due to a handoff consists of the handoff latency (T_H) and the PRP activation time (T_{ACT}). Another service disruption in SARAH may come from the ORP process. However, since SARAH performs the most of the ORP process simultaneously with ERP data forwarding through an extended reservation path, only a trivial delay is required for exchanging ORP messages and switching RSVP sessions. In our testbed, the ORP delay was measured to be about 8 ms, which does not significantly affect QoS guarantees. Our implementation performs an ORP process every time after an ERP process completes. This is done to reduce the waste of network resources due to the extended reservation path rather than to lighten the signalling overhead on BSs.

Figure 9 shows the average data rate variations measured on our testbed when SARAH and normal RSVP are applied. In our setup, the MH is initially located in BS2's cell as shown in Figure 7. Bandwidth of 250 kB (2000 kbps) has been reserved between a sender (CH) and a receiver (MH). The sender transmits 250 data packets per second with each packet size fixed at 1024 bytes. According to our measurements, the maximum capacity of the wired/wireless link in our testbed is about 9300/4700 kbps, respectively. Thus, background traffic of 9000 kbps was generated between the gateway router and BS1 so that the MH experiences congestion after moving from BS2 to BS1. Multi-Generator Tool (MGENT) [21] was used to generate the fixed-rate data traffic. Figure 9 shows that SARAH seamlessly maintains a stable data transmission rate even after the MH moves into the congested cell. Note that the momentary degradation of the data rate during a handoff is mainly caused by Mobile IP handoff latency. In contrast, RSVP does not guarantee seamless service quality after the MH moves into the congested network since it does not reserve resources in advance at the neighbouring BSs.

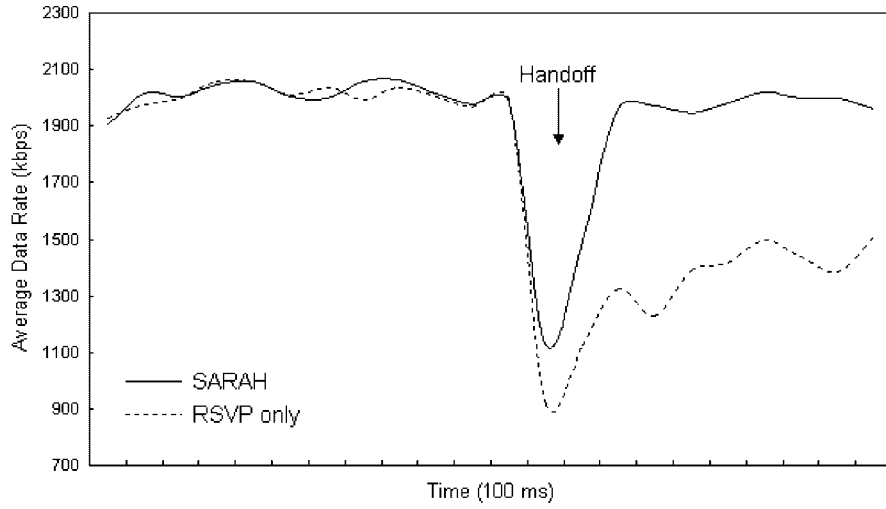


Figure 9. Average transmission rates.

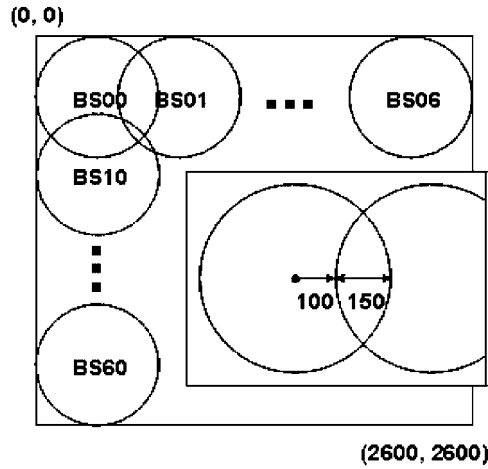


Figure 10. A 7×7 mesh simulation model.

6. SIMULATION STUDY

In addition to showing the operability of SARAH using a real implementation, the performance evaluation was carried out using the NS-2 network simulator [22]. Figure 10 shows the simulated network topology, which is a 7×7 mesh model where all the wireless cells have overlapped areas with their neighbours. As shown in the figure, all BSs are uniformly distributed over the entire test area. Each cell has a communication range of 250 m and the size of the overlapped area between two cells is 150 m. The L2 beacon interval for each BS is configured to be 100 ms. For simplicity, all BSs are designed to be 1-hop away from the gateway router. Figure 11 shows an example of a MH's movement history for our simulation, which follows the

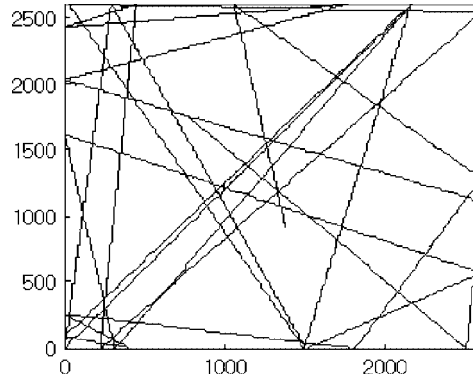


Figure 11. Random Direction Mobility Model.

Random Direction Mobility Model [23]. Each MH's initial location and direction is randomly chosen over the whole test area, and the direction of the movement is also randomly chosen whenever a MH arrives at the border of the test area. In our simulations, all MHs in the simulated area move at a speed of 3.5 m/s, which is a little faster than the speed of human walk.

We compared several QoS factors with the existing approaches, such as MRSVP [5, 6] and HMRSVP [10], to illustrate how strictly SARAH guarantees that a MH will complete its reservation session without any failures. The QoS factors compared in our simulation study are reservation blocking rate, reservation session loss rate, and reservation session completion rate [10]. The reservation blocking rate is the probability that an active reservation request in a wireless cell is blocked due to lack of network resources. The reservation session loss rate represents the probability for a MH to lose its reservation path after it handoffs to a new cell. The reservation session completion rate indicates the probability that a MH can maintain a reservation path without suffering from any reservation blocking and session loss until the reservation session finishes successfully. Therefore, this rate is a combinational effect of the reservation blocking rate and reservation session loss rate.

In order to compare the average advance reservation requirement (Q_P) for a MH of each approach, we first measured the average number of reachable BSs when a MH moves around the simulated area in Figure 10. To do this, we recorded 300 times the number of reachable BSs while a MH moves according to the random direction mobility pattern described above. The recorded values appeared to be about 1.49 on average. This value represents the total number of reservation requirements in SARAH, including active and pseudo reservations. When using MRSVP, HMRSVP, or SARAH, only one active reservation path is required for a MH. This means that Q_P for SARAH is about 0.49. On the other hand, the corresponding Q_P value for MRSVP is around 4. This difference is caused by that MRSVP always requires passive reservations at all neighbouring BSs (i.e. 4 neighbours in Figure 10) to prepare a MH's movement while SARAH establishes a PRP only when a MH finds a new reachable BS (i.e. when the MH receives a *beacon* frame from a BS that is not the currently associated one).

While the Q_P values for SARAH and MRSVP are hardly affected by the size of region (i.e. routing domain) and the network configuration, Q_P for HMRSVP varies with the proportion (P_B) of border cells to a region since HMRSVP establishes a passive reservation only for a MH that resides in a border cell. For inter-region handoff experiments, we designed the simulated

area shown in Figure 10 to be a region. Then P_B is calculated to be about 0.489 in 7×7 mesh network, and about 0.438 for 8×8 -mesh network, respectively. In the network topology shown in Figure 10, the proportion (P_O) of overlapping area in each cell is about 0.744. Thus, the average number of passive reservations (Q_P) for HMRSVP can be approximated as to be $(P_B \times P_O)/4$ since only one of 4 overlapping areas in a border cell is located between different regions. Q_P for HMRSVP is estimated to be around 0.091 in 7×7 mesh network.

Figure 12 shows the simulation result that measures the reservation blocking rates for the three schemes related to RSVP mobility support, namely, MRSVP, HMRSVP and SARAH. A parameter used for our simulation is the system offered load (ρ) in a wireless cell. The offered load is defined as the total amount of all active reservation requests from MHs in the simulation area. It can be modelled by four factors: Reservation inter-arrival time, reservation duration, total capacity of a wireless cell (C), and average number of MHs per each cell (N) [10]. When we assume the reservation inter-arrival time and the reservation duration follow exponential distributions with the mean $1/\lambda$ and $1/\mu$, respectively, the offered load is given by $\rho = N\lambda/C\mu$. In our simulation, we varied the offered load from 0.1 to 1 by increasing the number of MHs per each cell (N) while the other factors λ , μ and C are fixed to be constant values. As the offered load on the network increases (i.e. the number of simulated MHs increases), the average blocking rates increase in all the schemes under discussion. However, as can be seen in Figure 12, the reservation blocking rates of MRSVP are significantly higher than those of SARAH and HMRSVP. This is caused by the excessive reservation requirements at the neighbouring cells when MRSVP is applied. We can observe that the probability of reservation blocking is closely related to total number of excessive reservation requirements (Q_X) in each

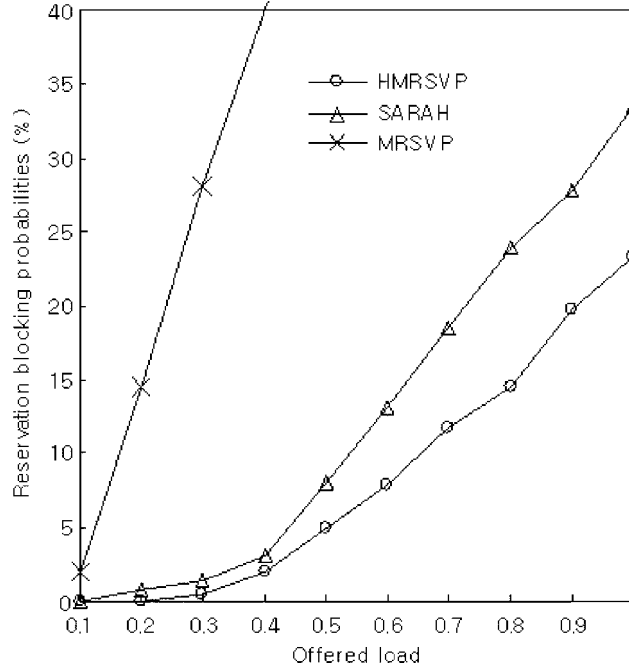


Figure 12. Reservation blocking rates.

scheme. Q_X can be estimated to be $\rho \times (Q_P + 1) - 1$ when $\rho \times (Q_P + 1)$ is greater than 1, otherwise Q_X is 0. The Q_X value for MRSVP is 4 when the offered load (ρ) is 1. On the other hand, HMRSVP has the lowest reservation blocking rate since it makes passive reservation only when a MH resides at a boundary cell of a region. For HMRSVP, Q_X appears to be only about 0.091 when ρ is 1. However, note that passive reservations for HMRSVP are converged on the border cells of the simulation region. This makes HMRSVP's reservation blocking rate in Figure 12 appear to be higher than the excessive amount of reservation requirements over network capacity.

Figure 13 shows the session loss rates for three schemes under discussion. While the resource-aware handoff direction scheme in SARAH does not affect the reservation blocking rate, the session loss rate of SARAH is considerably improved by employing the resource-aware handoff direction scheme (see the difference between SARAH and SARAH (without RAH) in Figure 13). This is caused by that the resource-aware handoff direction scheme increases the probability that a MH's handoff can be supported by a pre-established pseudo reservation. Without advance reservations, MHs will lose their reservation paths after they move into the congested cells. Even when the advanced reservation scheme is applied, a MH may also lose its reservation path when it moves into one of the cells that have rejected prior advance reservation requests. In general, more advance reservations give lower session loss rates. The number of advance reservations (Q_P) for MRSVP is 4, which is greater than those of SARAH and HMRSVP. However, the session loss rate of MRSVP is higher than those of SARAH and HMRSVP when the offered load in the network is low. This is because SARAH and HMRSVP do not make advance reservations at all the neighbouring cells, and thus more resources are

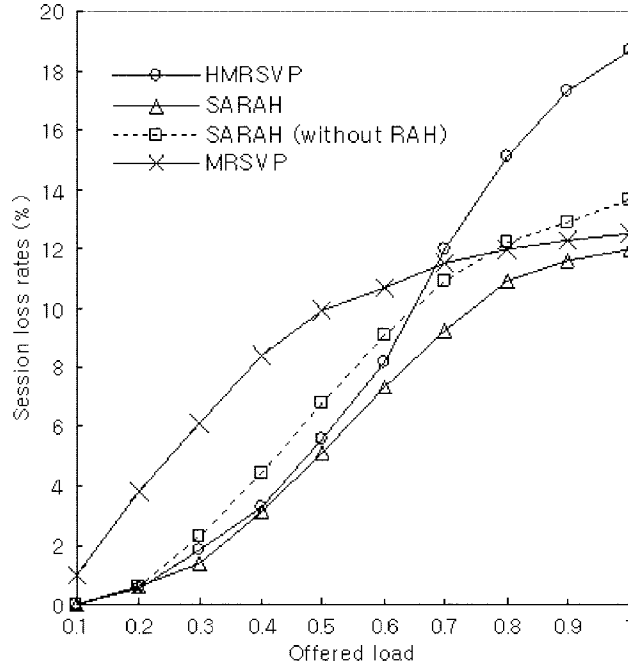


Figure 13. Reservation session loss rates.

available than MRSVP. When the offered load is higher than 0.7, HMRSVP has the highest session loss rate among all the schemes studied. Since HMRSVP does not make passive reservations for an intra-region handoff, it becomes easier for a MH to lose a reservation session after a movement as the network becomes congested. Figure 13 shows that SARAH provides the lowest session loss rate for all the discussed schemes. This is achieved by that SARAH suffers less from lack of resources due to excessive advance reservations than MRSVP and, moreover, it sufficiently makes advance reservations to support host mobility in the congested networks better than HMRSVP.

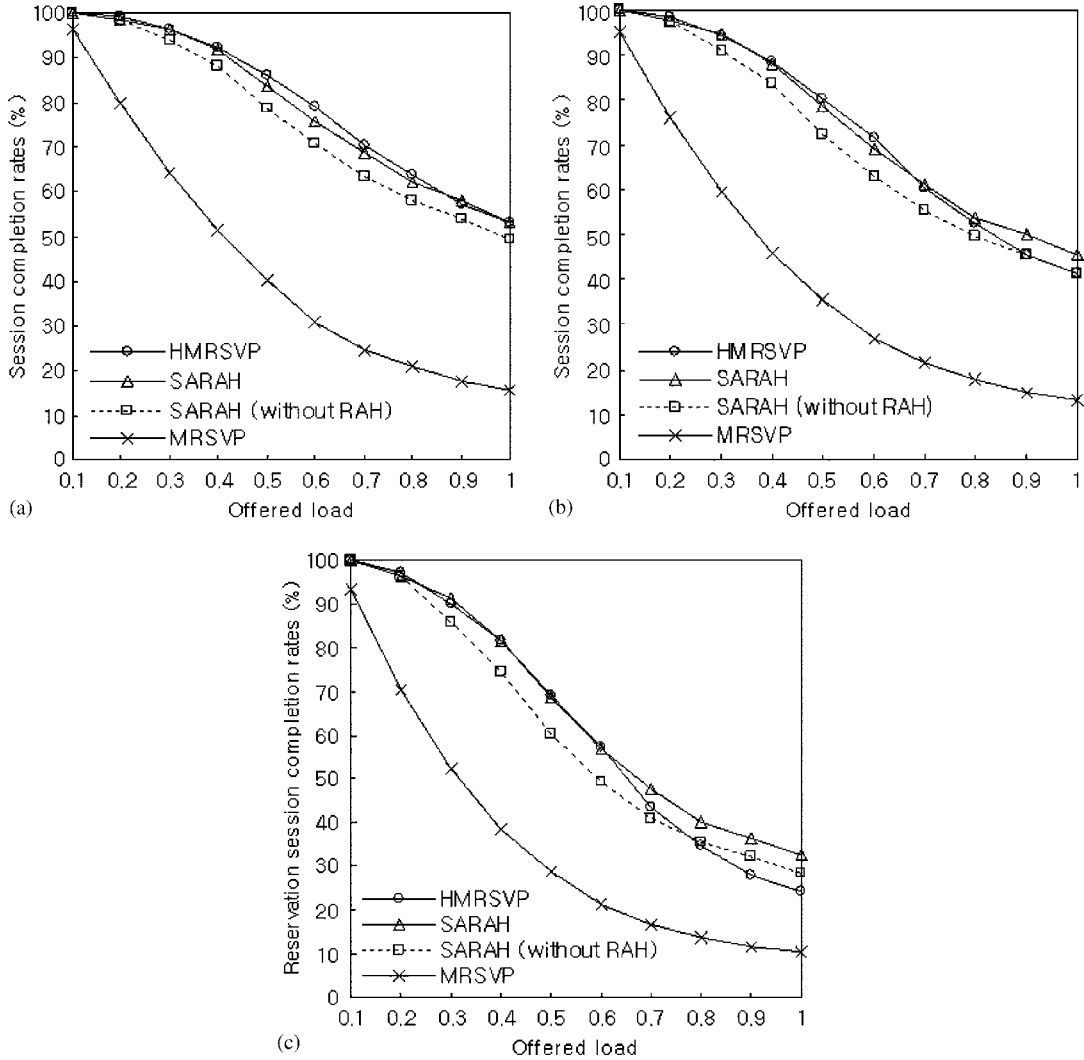


Figure 14. Reservation session completion rates: (a) MH speed: 3 m/s, $h \approx 4.78$; (b) MH speed: 5 m/s, $h \approx 3$; and (c) MH speed: 10 m/s, $h \approx 5.53$.

The reservation session completion rate is the most important evaluation factor since it directly reflects the grade of QoS guarantee for a MH. Thus, the session completion rate (C) can be estimated using the two aforementioned QoS factors, the reservation blocking rate and the session loss rate. It is given by $C = (1 - B)(1 - L)^N$, where B is the reservation blocking rate, L is the session loss rate, and N represents how many handoffs have occurred before the reservation session completes. This implies that the session completion rate becomes more affected by the session loss rate than the reservation blocking rate as the number of handoffs during a reservation session increases.

Figure 14 shows the simulation results that measured reservation session completion rates for the three schemes under discussion. Each graph in the figure illustrates a case that all MHs move in the simulation area according to the Random Direction Mobility Model [23] at a speed of 3, 5 and 10 m/s, respectively. Our simulations were performed for 3000 s while every MH was configured to complete its current reservation session and to restart a new one when the duration of the current session reaches 200 s. Since the reservation session duration is fixed to be a constant value, the average number of handoffs during a completed reservation session (h) is thoroughly affected by the speed of a MH. In our simulations, the h value for each case in Figure 14 appeared to be 1.78, 3.0 and 5.53, respectively.

As shown in Figure 14, MRSVP has the lowest session completion rate regardless of the speed of MHs. When the MH's speed is 3 m/s, the session completion rate for HMRSVP is slightly better than SARAH (see Figure 14(a)). This difference is mainly caused by that HMRSVP has a lower reservation blocking rate than SARAH. However, when the MH's speed is 5 m/s (Figure 14(b)), the session completion rate of SARAH becomes similar with that of HMRSVP. In this case, SARAH outperforms HMRSVP as the offered load increases to be higher than 0.7, i.e. the network is highly congested. The reason for this is the session loss rate of HMRSVP increases rapidly as the offered load increases as shown in Figure 13. Figure 14(c) shows that the reservation session completion rate of SARAH is always better than that of HMRSVP when the speed of MHs becomes greater than 10 m/s. This shows that, when the networks is congested and the mobility of host is high, the proposed SARAH mechanism provides relatively better performance than the existing approaches such as HMRSVP.

7. CONCLUSIONS

This paper proposed a new mechanism, called SARAH, which guarantees seamless QoS support for a host moving in the mobile Internet. To support QoS-guaranteed handoffs, SARAH extends a reservation path by activating an advance reservation, called pseudo reservation path (PRP), and attaching it to the original reservation path. It also dynamically optimizes the extended reservation path to avoid the infinite path extension problem.

SARAH addresses the excessive reservation requirements due to establishment of multiple advance reservations. It significantly reduces the number of required PRPs with the movement detection scheme using layer-2 (link-layer) functionalities. Another novel feature of SARAH is the resource-aware handoff direction scheme to manage the network resources more efficiently. With this scheme, a MH chooses its next BS according to not only the signal strength of layer 2 control frames, but also the available resources in the reachable BS. This considerably increases the probability that a MH can successfully complete a reservation session without suffering from any QoS degradation. In addition, the proposed mechanism requires fewer functional and

structural changes in the existing network components, and imposes no modification or enhancement on the existing RSVP and Mobile IP protocol. It also integrates all the enhanced features for pseudo reservation and path extension only into the leaf BSs and MHs.

Our experimental results demonstrate that SARAH significantly saves network resources required for the establishment of PRPs without degrading the QoS guarantees. The performance comparison with the existing approaches, such as MRSVP and HMRSVP, shows that SARAH is a more efficient way to guarantee QoS in the congested networks that the host mobility is high. This becomes an important advantage of SARAH as well as requiring fewer changes to the existing Internet architecture. As a future work, the performance improvement of SARAH due to reservation load balancing will be studied.

ACKNOWLEDGEMENTS

This paper was supported in part by the GRID Middleware Center in OITRC.

REFERENCES

1. Branden R, Zhang L, Berson S, Herzog S, Jamin S. Resource ReSerVation Protocol (RSVP)—Version 1 functional specification. *RFC 2205, IETF*, September 1997.
2. Blake S, Black D, Carlson M, Davies E, Wang Z, Weiss W. An architecture for differentiated services. *RFC 2475 on IETF*, December 1998.
3. Perkins CE. IP mobility support. *RFC 2002 on IETF*, October 1996.
4. Terzis A, Srivastava M, Zhang L. A simple QoS signaling protocol for mobile hosts in the integrated service internet. *18th Annual Joint Conference on the IEEE Computer and Communications Societies, INFOCOM 99*, vol. 3. New York, NY, U.S.A., March 1999; 1011–1018.
5. Talukdar AK, Badrinath BR, Acharya A. MRSVP: a resource reservation protocol for an integrated service networks with mobile hosts. *ACM Wireless Networks* 2001; 7(1):5–19.
6. Talukdar AK, Badrinath BR, Acharya A. On accommodating mobile hosts in an integrated services packet network. *16th Annual Joint Conference on the IEEE Computer and Communications Societies, INFOCOM 97*, vol. 3. Kobe, Japan, April 1997; 1046–1053.
7. Pasklis S, Kaloxylas A, Zervas E, Merakos L. An efficient RSVP-mobile IP interworking scheme. *ACM Mobile Networks and Applications* 2001; 8(3):630–637.
8. Chen W, Huang L. RSVP mobility support: a signaling protocol for integrated services internet with mobile hosts. *19th Annual Joint Conference on the IEEE Computer and Communications Societies, INFOCOM 2000*, vol. 3. Tel-Aviv, Israel, March 2000; 1283–1292.
9. Mahadevan I, Sivalingam KM. Architecture and experimental results for quality of service in mobile networks using RSVP and CBQ. *ACM Wireless Networks* 2000; 6(3):221–234.
10. Tseng C, Lee G, Liu R. HMRSVP: a hierarchical mobile RSVP protocol. *ACM Wireless Networks* 2003; 9(2):95–102.
11. Gustafsson E, Jonson A, Perkins CE. Mobile IP regional registration. *Internet Draft on IETF*, October 2002.
12. Malki KE, Calhoun PR, Hiller T, Kempf J, McCann PJ, Singh A, Soliman H, Thalanany S. Low latency handoffs in mobile IPv4. *Internet Draft on IETF*, June 2002.
13. Calhoun P. FA assisted hand-off. *Internet Draft on IETF*, March 2000.
14. Wireless LAN Working Group, Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. ISO/IEC 8802.11:1999(E), IEEE Std 802.11 1999 Edition, August 1999.
15. Lucent Technologies Inc., Roaming with WaveLAN/IEEE 802.11. *WaveLAN Technical Bulletin 021/A*, December 1998.
16. Fenner W. Internet group management protocol, Version 2. *RFC 2236 on IETF*, November 1997.
17. WaveLAN, <http://www.agere.com/client/wlan.html>
18. ALTQ: Alternate Queueing, <http://www.csl.sony.co.jp/person/kjc/kjc/software.html>
19. Dynamics—HUT Mobile IP, <http://www.cs.hut.fi/Research/Dynamics>
20. RSVP Code rel4.2a3, <ftp://ftp.isi.edu/rsvp/release/>
21. MGEN: The Multi-Generator Tool, <http://manimac.itd.nrl.navy.mil/MGEN/>
22. The Network simulator—NS-2, <http://www.isi.edu/nsnam/ns/>
23. Camp T, Boleng J, Davies V. A survey of mobility models for ad hoc network research. *Wireless Communications and Mobile Computing* 2002; 2(5):483–502.